



HACK YOUR DATA BEAUTIFUL: WELCOME!

DALE BARR

WHY WRITE CODE? CODE IS...

- **portable** : plain text is easy to share, non-proprietary
- **transparent** : choices are made explicit
- **self-documenting** : no need to separately log actions
- **empowering** : enhances efficiency and builds confidence

THINGS I WISH I HAD KNOWN:

There is no “geek gene”

Evidence That Computer Science Grades Are Not Bimodal

Elizabeth Patitsas, Jesse Berlin, Michelle Craig, and Steve Easterbrook
Department of Computer Science
University of Toronto
Toronto, Ontario, Canada
patitsas,mcraig,sme@cs.toronto.edu and jesse.berlin@mail.utoronto.ca

ABSTRACT

Although it has never been rigourously demonstrated, there is a common belief that CS grades are bimodal. We statistically analyzed 778 distributions of final course grades from a large research university, and found only 5.8% of the distributions passed tests of multimodality. We then devised a psychology experiment to understand why CS educators believe their grades to be bimodal. We showed 53 CS professors a series of histograms displaying ambiguous distributions and asked them to categorize the distributions. A random half of participants were primed to think about the fact that CS grades are commonly thought to be bimodal; these participants were more likely to label ambiguous distributions as “bimodal”. Participants were also more likely to label distributions as bimodal if they believed that some students are innately predisposed to do better at CS. These results suggest that bimodal grades are instructional folklore in CS, caused by confirmation bias and instructor beliefs about their students.

inform their practice [13], and these beliefs may or may not be based on empirical evidence.

1.1 Explanations of Bimodality

A number of explanations have been presented for why CS grades are bimodal, all of which begin with the assumption that this is the case.

1.1.1 Prior Experience

A bimodal distribution generally indicates that two distinct populations have been sampled together [5]. One explanation for bimodal grades is that CS1 classes have two populations of students: those with experience, and those without it [1].

High school CS is not common in many countries, and so students enter university CS with a range of prior experience. However, this explanation fits students into two bins. Prior experience is not as simple as “have it” vs. not – there is a large range on how much prior experience students

THINGS I WISH I HAD KNOWN:

You can't break your data!

- R is a functional language that operates on data *in memory*.
- Each transformation creates a new copy of your data.
- As long as you avoid file operations, the original data remains untouched by actions in your script.

THINGS I WISH I HAD KNOWN

The greatest beneficiary of working reproducibly is *future you*

Your primary collaborator is yourself 6 months from now, and your past self doesn't answer emails

-Software Carpentry

...and now is the best time to invest in that person

THINGS I WISH I HAD KNOWN

You will never stop learning

As of today, you are a programmer! Congratulations!

CONSOLIDATE YOUR SKILLS, AND PASS IT ON

- Stay engaged!
- Build local communities
 - journal clubs, user groups, online messaging (slack)
- Join extended communities on Twitter
 - #rstats, @R4DScommunity, @RLadiesGlobal
 - Tidy Tuesday: github.com/rfordatascience/tidytuesday

HAVE FUN!